

# Modelado y visualización de relaciones entre contaminantes del aire y salud pública

Selene Berenice Prado Prado  
Ing. en Tecnología de Software

Universidad Autónoma de Nuevo León  
Facultad de Ingeniería Mecánica y Eléctrica

8 de abril de 2022

# Índice

## 1 Introducción

- Motivación
- Hipótesis
- Objetivos

## 2 Antecedentes

- Monitoreo de calidad del aire
- Series de tiempo
- Clasificación de enfermedades
- Regresión lineal

- Regresión lineal múltiple

## 3 Estado del arte

## 4 Solución propuesta

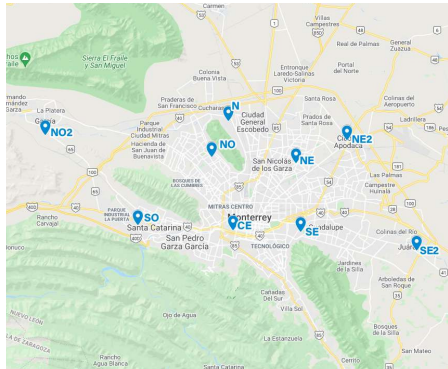
- Herramientas
- Implementación de la solución

## 5 Experimentos

- Niveles de PM10
- Niveles de PM2.5

## 6 Conclusiones

# Introducción



**Figura 1:** Estaciones de monitoreo de la calidad del aire pertenecientes al Sistema Integral de Monitoreo Ambiental (SIMA) [1].

# Motivación

## Motivación

Con el presente trabajo se busca aportar a la creación de nuevas herramientas que permitan observar y estudiar dichas relaciones con el fin de ayudar a tomar medidas adecuadas que permitan aminorar los efectos negativos de los contaminantes del aire en la salud.

# Hipótesis

## Hipótesis

Los modelos de regresión permiten obtener gráficos donde se pueden observar las relaciones entre el número de ingresos hospitalarios y los niveles de contaminantes del aire.

# Objetivos

## Objetivo general

Generar, implementar y evaluar modelos que muestran las relaciones existentes entre contaminantes del aire y salud pública tiene la finalidad de apoyar a la implementación de estrategias que aminoran los efectos negativos de los contaminantes del aire en la salud de las personas. Con los modelos generados se puede tener una herramienta que permite identificar gráficamente las relaciones con solo proporcionarle el conjunto de datos.

# Monitoreo de calidad del aire

## Contaminantes del aire

- Monóxido de carbono (CO).
- Dióxido de azufre (SO<sub>2</sub>).
- Óxidos de nitrógeno (NO<sub>x</sub>).
- Ozono (O<sub>3</sub>).
- Partículas de tamaño menor a 10 micrómetros (PM10).
- Partículas de tamaño menor a 2.5 micrómetros (PM2.5).

# Series de tiempo

## Definición de serie de tiempo

Las series de tiempo se pueden definir como un conjunto de observaciones tomadas en un tiempo  $t$  determinado. Los estudios de series de tiempo relacionan estadísticamente los cambios temporales en la repercusión de cambios en la concentración de un contaminante en la población [2].



# Serie de tiempo

## Definición de semana epidemiológica

Una semana epidemiológica es un estándar de medición temporal que se utiliza para comparar datos en ventanas de tiempo definidas. La primera semana epidemiológica del año termina el primer sábado de enero de cada año [3].

# Clasificación de enfermedades

## Clasificación Internacional de Enfermedades (CIE)

Existe un instrumento estadístico y sanitario para identificar enfermedades llamado Clasificación Internacional de Enfermedades (CIE) que agrupa enfermedades en epidémicas, generales, locales ordenadas por origen geográfico, trastornos del desarrollo y lesiones [4].

# Regresión lineal

## Definición de regresión lineal

La tendencia  $w_0$  de una serie de tiempo puede ser obtenida a partir de una regresión lineal de la misma [5]. Una regresión lineal es una metodología inferencial supervisada que busca predecir valores y dado un vector de variables de entrada  $t$  por medio del ajuste de coeficientes  $w$  de la función lineal:

$$\hat{y}(t, w) = w_0 + w_1x_1 + \dots + w_tx_t. \quad (1)$$

# Regresión lineal múltiple

## Definición de regresión lineal múltiple

Un modelo de regresión múltiple es un modelo complemento de la regresión lineal simple, el cual tiene dos o más variables independientes  $k$  que pueden influir en una variable dependiente  $y$ . Peniche-Camps and Cortez-Huerta [6] expresan la regresión múltiple mediante la siguiente ecuación:

$$y = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k + \varepsilon. \quad (2)$$

## Estado del arte

**Cuadro 1:** Comparación de trabajos frente al desarrollado, donde ✓ indica que cumple con esta característica y × no cumple con esta característica.

Trabajo	Modelos de regresión lineal	Modelos de predicción	Evaluación de modelos	Estudio de contaminantes del aire	Estudio de problemas de salud
Martin and Bayle [7]	✓	×	×	✓	✓
Guarnaccia et al. [8]	×	✓	✓	✓	×
Julia et al. [9]	✓	✓	×	✓	✓
Zhang et al. [10]	✓	✓	×	✓	✓
Kim et al. [11]	✓	×	×	✓	✓
Liu et al. [12]	×	✓	✓	✓	✓
To et al. [13]	×	×	×	✓	✓
Breton et al. [14]	✓	✓	×	✓	✓
Gupta et al. [15]	✓	✓	×	✓	✓
El presente trabajo	✓	✓	✓	✓	✓

# Solución propuesta

## Solución propuesta

La solución propuesta se compone de cuatro fases principales: **recolección de datos, selección y agrupación de datos, visualización de la evolución de las variables e implementación de modelos.**

# Herramientas

Cuadro 2: Herramientas utilizadas.

Herramienta	Versión	URL
Python	3.8.8	<a href="https://www.python.org/">https://www.python.org/</a>
Jupyter Notebook	6.3.0	<a href="https://jupyter.org/">https://jupyter.org/</a>
Imageio	2.9.0	<a href="https://imageio.readthedocs.io/">https://imageio.readthedocs.io/</a>
LatexTable	0.2.1	<a href="https://pypi.org/project/latexTable/">https://pypi.org/project/latexTable/</a>
Matplotlib	3.3.4	<a href="https://matplotlib.org/">https://matplotlib.org/</a>
NumPy	1.20.1	<a href="http://www.numpy.org/">http://www.numpy.org/</a>
Pandas	1.2.4	<a href="https://pandas.pydata.org/">https://pandas.pydata.org/</a>
Seaborn	0.11.1	<a href="https://seaborn.pydata.org/">https://seaborn.pydata.org/</a>
Scikit-learn	0.24.1	<a href="https://scikit-learn.org/">https://scikit-learn.org/</a>
SciPy	1.6.2	<a href="https://docs.scipy.org/">https://docs.scipy.org/</a>
Statsmodels	0.12.2	<a href="https://www.statsmodels.org/">https://www.statsmodels.org/</a>
Texttable	1.6.4	<a href="https://pypi.org/project/texttable/">https://pypi.org/project/texttable/</a>

# Implementación de la solución

## Visualizaciones generadas

Los tipos de visualizaciones generadas son:

- Series de tiempo.
- Gráficos de radar.

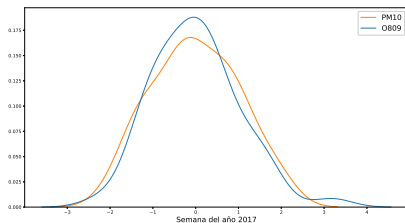
## Modelos generados

Los tipos de modelos generados son:

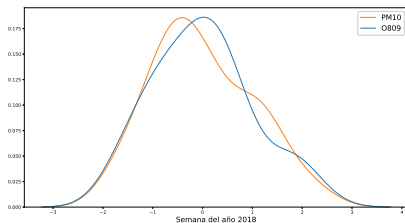
- Modelos de regresión lineal.
- Modelos de regresión lineal múltiple.



## Niveles de PM10



**Figura 2:** Evolución de los niveles de PM10 y el número de egresos con la CIE O809 en el 2017.



**Figura 3:** Evolución de los niveles de PM10 y el número de egresos con la CIE O809 en el 2018.



# PM10 2017

**Cuadro 4:** Resultados regresión lineal múltiple PM10 2017

<b>Variable Dep.:</b>	y	<b>R<sup>2</sup>:</b>	0.313			
<b>Modelo:</b>	OLS	<b>Método:</b>	Mínimos cuadrados			
<b>Error:</b>	0.226					
	<b>coef</b>	<b>std err</b>	<b>t</b>	<b>P&gt;  t </b>	<b>[0.025</b>	<b>0.975]</b>
<b>const</b>	0.5859	0.146	4.002	0.000	0.289	0.883
<b>O809</b>	-0.3019	0.122	-2.472	0.018	-0.550	-0.054
<b>O829</b>	0.2685	0.123	2.185	0.036	0.019	0.518
<b>O759</b>	-0.2229	0.182	-1.222	0.230	-0.593	0.147
<b>O069</b>	-0.1441	0.120	-1.200	0.238	-0.388	0.100
<b>K802</b>	-0.0456	0.133	-0.344	0.733	-0.315	0.224



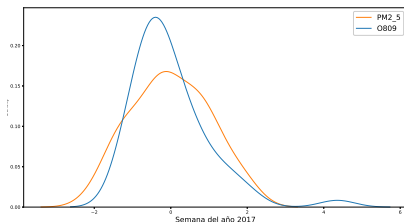
# PM10 2018

**Cuadro 6:** Resultados regresión lineal múltiple PM10 2018

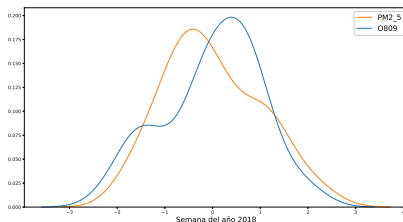
**Variable Dep.:**  $y$  **R<sup>2</sup>:** 0.182  
**Modelo:** OLS **Método:** Mínimos cuadrados  
**Error:** 0.159

	<b>coef</b>	<b>std err</b>	<b>t</b>	<b>P &gt;  t </b>	<b>[0.025</b>	<b>0.975]</b>
<b>const</b>	0.3885	0.210	1.849	0.073	-0.038	0.815
<b>O809</b>	-0.0114	0.188	-0.061	0.952	-0.392	0.370
<b>O829</b>	0.0854	0.214	0.400	0.692	-0.349	0.520
<b>K802</b>	0.3088	0.167	1.852	0.072	-0.030	0.647
<b>O342</b>	-0.1708	0.208	-0.820	0.418	-0.593	0.252
<b>N40X</b>	-0.1239	0.167	-0.740	0.464	-0.464	0.216

## Niveles de PM2.5



**Figura 4:** Evolución de los niveles de PM2.5 y el número de egresos con la CIE O809 en el 2017.



**Figura 5:** Evolución de los niveles de PM2.5 y el número de egresos con la CIE O809 en el 2018.



# PM2.5 2017

**Cuadro 8:** Resultados regresión lineal múltiple PM2.5 2017

<b>Variable Dep.:</b>	y	<b>R<sup>2</sup>:</b>	0.210			
<b>Modelo:</b>	OLS	<b>Método:</b>	Mínimos cuadrados			
<b>Error:</b>	0.175					
	<b>coef</b>	<b>std err</b>	<b>t</b>	<b>P&gt;  t </b>	<b>[0.025</b>	<b>0.975]</b>
<b>const</b>	0.5345	0.158	3.373	0.002	0.213	0.856
<b>O809</b>	-0.1754	0.132	-1.327	0.193	-0.444	0.093
<b>O829</b>	0.0701	0.133	0.527	0.602	-0.200	0.340
<b>O759</b>	-0.2585	0.197	-1.309	0.199	-0.659	0.142
<b>O069</b>	-0.2148	0.130	-1.652	0.108	-0.479	0.049
<b>K802</b>	-0.1164	0.144	-0.811	0.423	-0.408	0.175



# PM2.5 2018

**Cuadro 9:** Resultados obtenidos PM2.5 2018

CIE	$\rho$	$R^2$	Valor $p$	$\epsilon$
O809	-0.354	0.057	0.133	0.259
O829	0.225	0.046	0.177	0.256
K802	0.273	0.089	0.058	0.159
O342	-0.443	0.149	0.013	0.245
N40X	-0.074	0.001	0.842	0.241

$\rho$  = Coeficiente de correlación de Pearson

$\epsilon$  = RMSE para medir el error

# PM2.5 2018

**Cuadro 10:** Resultados regresión lineal múltiple PM2.5 2018

**Variable Dep.:**  $y$       **R<sup>2</sup>:** 0.259  
**Modelo:** OLS      **Método:** Mínimos cuadrados  
**Error:** 0.174

	<b>coef</b>	<b>std err</b>	<b>t</b>	<b>P &gt;  t </b>	<b>[0.025</b>	<b>0.975]</b>
<b>const</b>	0.7042	0.193	3.652	0.001	0.313	1.096
<b>O809</b>	-0.0863	0.172	-0.501	0.620	-0.436	0.264
<b>O829</b>	-0.1084	0.196	-0.553	0.584	-0.507	0.290
<b>K802</b>	0.3006	0.153	1.964	0.058	-0.010	0.611
<b>O342</b>	-0.3311	0.191	-1.733	0.092	-0.719	0.057
<b>N40X</b>	-0.2053	0.154	-1.336	0.190	-0.517	0.107

# Conclusiones

## Conclusión

En las series de tiempo se observa que la cantidad de egresos de la mayoría de las CIE estudiadas presentan una línea de evolución similar al contaminante PM10. Además, se observó que para el estudio de las relaciones entre los contaminantes y las CIE se obtuvieron mejores resultados con los modelos de regresión lineal múltiple frente a los modelos de regresión lineal simple.

# Conclusiones

## Trabajo a futuro

- Recolección de más datos de los niveles de contaminantes y de egresos para el estudio de años más recientes.
- Realizar un estudio de las relaciones de los niveles de determinados contaminantes y la cantidad de egresos por CIE empleando modelos de regresión lineal múltiple.
- Creación de un mapa interactivo donde se pueda observar por región los niveles de los contaminantes y la cantidad de egresos.

# Referencias I

- [1] SIMA. Sistema integral de monitoreo ambiental, 2015. URL <http://aire.nl.gob.mx>.
- [2] Peter J Brockwell, Peter J Brockwell, Richard A Davis, and Richard A Davis. *Introduction to time series and forecasting*. Springer, 2002. ISBN 9780387216577.
- [3] Jorge R Arias. What is an epidemiological week and why do we use them. *Skeeter*, 66(1):7, 2006.
- [4] World Health Organization et al. International classification of diseases. 2016. *World Health Organization*, 2016. URL <https://icd.who.int/browse10/2016/en>.
- [5] Richard B Darlington and Andrew F Hayes. *Regression analysis and linear models: Concepts, applications, and implementation*. Guilford Publications, 2016. ISBN 9781462521135.
- [6] Salvador Peniche-Camps and Mauro Cortez-Huerta. La costumbre al envenenamiento: El caso de los contaminantes atmosféricos de la ciudad de guadalajara, méxico. *Revista de Ciencias Ambientales*, 54(2): 1–19, 2020. doi:10.15359/rca.54-2.1.
- [7] Raquel Martín Martín and Marciano Sánchez Bayle. Impacto de la contaminación ambiental en las consultas pediátricas de atención primaria: estudio ecológico. In *Anales de Pediatría*, volume 2, pages 80–85. Elsevier, 2018. doi:10.1016/j.anpedi.2017.06.013.
- [8] Claudio Guarnaccia, Julia Griselda Ceron Breton, Rosa Maria Ceron Breton, Carmine Tepedino, Joseph Quartieri, and Nikos E Mastorakis. Arima models application to air pollution data in monterrey, mexico. In *AIP Conference Proceedings*, volume 1, pages 1–5. AIP Publishing LLC, 2018. doi:10.1063/1.5045447.
- [9] Ana Julia, Diana A FC Lichtenfels, Kim Van Der Plaats, Cleo C de Jong, Dirkje S van Diemen, Ivana Postma, Cornelia M Nedeljkovic, Najaf van Duijn, Sacha Amin, Maaike la Bastide-Van Gemert, De Vries, et al. Long-term air pollution exposure, genome-wide dna methylation and lung function in the lifelines cohort study. *Environmental health perspectives*, 126(2):027004, 2018. doi:10.1289/EHP2045.

## Referencias II

- [10] Zilong Zhang, Bin Dong, Shanshan Li, Gongbo Chen, Zhaogeng Yang, Yanhui Dong, Zhenghe Wang, Jun Ma, and Yuming Guo. Exposure to ambient particulate matter air pollution, blood pressure and hypertension in children and adolescents: a national cross-sectional study in china. *Environment international*, 128: 103–108, 2019. doi:10.1016/j.envint.2019.04.036.
- [11] Jeniffer S Kim, Zhonghua Chen, Tanya L Alderete, Claudia Toledo-Corral, Fred Lurmann, Kiros Berhane, and Frank D Gilliland. Associations of air pollution, obesity and cardiometabolic health in young adults: The meta-air study. *Environment international*, 133:105–180, 2019. doi:10.1016/j.envint.2019.105180.
- [12] Yuewei Liu, Jingju Pan, Hai Zhang, Chunxiang Shi, Guo Li, Zhe Peng, Jixuan Ma, Yun Zhou, and Lan Zhang. Short-term exposure to ambient air pollution and asthma mortality. *American journal of respiratory and critical care medicine*, 200(1):24–32, 2019. doi:10.1164/rccm.201810-1823OC.
- [13] Teresa To, Jingqin Zhu, Dave Stieb, Natasha Gray, Ivy Fong, Lauren Pinault, Michael Jerrett, Alain Robichaud, Richard Ménard, Aaron van Donkelaar, et al. Early life exposure to air pollution and incidence of childhood asthma, allergic rhinitis and eczema. *European Respiratory Journal*, 55(2), 2020. doi:10.1183/13993003.00913-2019.
- [14] Rosa Maria Cerón Breton, Julia Céron Breton, María de la Luz Espinosa Fuentes, Jonathan Kahl, Alberto Antonio Espinosa Guzman, Rocío García Martínez, Claudio Guarnaccia, Reyna del Carmen Lara Severino, Evangelina Ramirez Lara, and Antonella Bianca Francavilla. Short-term associations between morbidity and air pollution in metropolitan area of monterrey, mexico. *Atmosphere*, 12(10):1352, 2021. doi:10.3390/atmos12101352.
- [15] Ankit Gupta, Hemant Bherwani, Sneha Gautam, Saima Anjum, Kavya Musugu, Narendra Kumar, Avneesh Anshul, and Rakesh Kumar. Air pollution aggravating covid-19 lethality? exploration in asian cities using statistical models. *Environment, Development and Sustainability*, 23(4):6408–6417, 2021. doi:10.1007/s10668-020-00878-9.